

RESEARCH ARTICLE

On the Origins and Control of Community Types in the Human Microbiome

Travis E. Gibson ¹, Amir Bashan ¹, Hong-Tai Cao ^{1,2,3}, Scott T. Weiss ¹, Yang-Yu Liu ^{1,4*}

1 Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Department of Electrical Engineering, University of Southern California, Los Angeles, California, United States of America, **3** Chu Kochen Honors College, College of Electrical Engineering, Zhejiang University, Hangzhou, Zhejiang, China, **4** Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America

☯ These authors contributed equally to this work.

* yyl@channing.harvard.edu



Abstract

Microbiome-based stratification of healthy individuals into compositional categories, referred to as “enterotypes” or “community types”, holds promise for drastically improving personalized medicine. Despite this potential, the existence of community types and the degree of their distinctness have been highly debated. Here we adopted a dynamic systems approach and found that heterogeneity in the interspecific interactions or the presence of strongly interacting species is sufficient to explain community types, independent of the topology of the underlying ecological network. By controlling the presence or absence of these strongly interacting species we can steer the microbial ecosystem to any desired community type. This open-loop control strategy still holds even when the community types are not distinct but appear as dense regions within a continuous gradient. This finding can be used to develop viable therapeutic strategies for shifting the microbial composition to a healthy configuration.

OPEN ACCESS

Citation: Gibson TE, Bashan A, Cao H-T, Weiss ST, Liu Y-Y (2016) On the Origins and Control of Community Types in the Human Microbiome. *PLoS Comput Biol* 12(2): e1004688. doi:10.1371/journal.pcbi.1004688

Editor: Jeff Gore, MIT, UNITED STATES

Received: September 11, 2015

Accepted: December 1, 2015

Published: February 11, 2016

Copyright: © 2016 Gibson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data used in this study is publicly available from the metagenomics analysis server MG-RAST:4457768.3-4459735.3 and can also be accessed from Qiita (<http://qiita.ucsd.edu>) under study ID 550. The processed data was downloaded as biom file “67_otu_table.biom” (2014-11-17 13:18:50.591389).

Funding: STW was supported by the National Institutes of Health (<http://www.nih.gov>) under grant number R01 HL091528 and TEG, YYL and AB were supported by the John Templeton Foundation (<https://www.templeton.org>) under grant number (51977). The funders had no role in study design, data

Author Summary

We coexist with a vast number of microbes that live in and on our bodies, and play important roles in physiology and disease. Two interesting phenomena have been observed in the human microbiome. The first is the stratification of healthy individuals based on the relative abundances of their microbes, which holds promise for drastically improving personalized medicine. The second is the astounding success of fecal microbial transplantation in treating certain diseases related to disordered microbiomes. Surprisingly, both phenomena have not been analytically or quantitatively understood, despite a few early qualitative attempts. This work shows that through a dynamic systems and control theoretical approach the success of fecal microbial transplantation can be explained and that the microbiome-based stratification can be as simple as the existence of strongly interacting species.

collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Rather than simple passengers in and on our bodies, commensal microorganisms have been shown to play key roles in our physiology and in the evolution of several chronic diseases [1, 2]. Many scientific advances have been made through the work of large-scale, consortium-driven metagenomic projects, such as the *Metagenomics of the Human Intestinal Tract* (Meta-HIT) [3] and the *Human Microbiome Project* (HMP) [4, 5]. In particular, the HMP has analyzed the largest cohort and set of distinct, clinically relevant body habitats to date, in order to characterize the ecology of human-associated microbial communities [4]. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the translational applications of the human microbiome.

A recent study proposed that a healthy gut microbiome falls within one of three distinct community types, which the authors coined as “enterotypes” [6]. More specifically, the authors calculated the relative abundance profiles of microbiota at the genus level and then performed standard cluster analysis, finding three distinct clusters (enterotypes). Each enterotype is a dominated by a particular genus (*Bacteroides*, *Prevotella*, or *Ruminococcus*) but not affected by gender, age, body mass index, or nationality of the host. These results suggest that enterotyping could be an efficient way to stratify healthy human individuals. The development of personalized microbiome-based therapies would then simplify to shifting an unhealthy microbiome to one of the distinct healthy configurations.

A meta-analysis, however, suggested that enterotypes, or in general community types, could be an artifact of the small sample size in [6] and what one should expect is a continuous gradient with dense regions rather than distinct clusters [7]. The level of discreteness or continuity of the community types remains unclear. Interestingly, samples in the dense regions of this gradient are either highly abundant or deficient in *Bacteroides* [7], indicating that community types could still emerge as the dense regions within a continuous gradient. Indeed, some recent work actually supports the notion of distinct community types [8–12].

We still lack consensus on the nature and origins of community types [13–17]. In principle the presence of community types could be explained by several different mechanisms. For instance, there may be true multi-stability, i.e. multiple stable states with the same set of microbial species present in the same environment [18]. Although this type of multi-stability has been well discussed in macro-ecological systems [19], its detection in host-associated microbial communities is rather difficult and has not been demonstrated experimentally [15]. Host heterogeneity is another possible mechanism, leading to host-specific microbial dynamics (parameterized by host-specific intra- and inter-species interactions). If those interactions, which serve as parameters of the host-associated microbial ecosystems, can be classified into distinct groups, then we can numerically demonstrate that distinct community types will naturally emerge (S1 Text Sec. 6.2 and 7.1). However, the presence of classifiable microbial dynamics has not been experimentally detected. Moreover, the overwhelming success of *Fecal Microbiota Transplantation* (FMT) in treating *recurrent Clostridium Difficile Infection* (rCDI) suggests that host heterogeneity is likely playing a minor role in terms of its effect on intra- and inter-species interactions [20–22].

Here we proposed a simple mechanism, without assuming multi-stability or host heterogeneity, to explain the origins of community types. In particular, using a dynamic systems approach, we studied compositional shift as a function of species collection and demonstrated that with heterogeneous interspecific interactions, a phenomenon often observed in macroecology [23–25], community types can naturally emerge. Interestingly, this result is independent of the topology of the underlying ecological network. To our knowledge, this is the first

quantitative attempt to explore the analytical basis of community types. Furthermore, community types, even when they weakly exist, can be manipulated efficiently by controlling the *Strongly Interacting Species* (SISs) only. This provides theoretical justification for translational applications of the human microbiome. Note that in this paper we use the term species in the general context of ecology, i.e. a set of organisms adapted to a particular set of resources in the environment, rather than the lowest taxonomic rank. One could think of organizing microbes by genus or operational taxonomical units as well.

Dynamic Model

The human microbiome is a complex and dynamic ecosystem [26]. When modeling a dynamic system we should first decide how complex the model needs to be so as to capture the phenomenon of interest. A detailed model of the intestinal microbiome would include mechanistic interactions among cells, spatial structure of the human intestinal tract, as well as host-microbiome interactions [27–30]. That level of detail however is not necessary for this study, because we are primarily interested in exploring the impact that any given species has on the abundance of other species. To achieve that, a population dynamics model such as the canonical *Generalized Lotka-Volterra* (GLV) model is sufficient [15, 31]. Indeed, GLV dynamics leveraging current metagenome data has already been used for predictive modeling of the intestinal microbiota [32–34]. Consider a collection of n species in a habitat with the population of species i at time t denoted as $x_i(t)$. The GLV model assumes that the species populations follow a set of ordinary differential equations

$$\dot{x}_i(t) = r_i x_i(t) + x_i(t) \sum_{j=1}^n a_{ij} x_j(t), \quad i = 1, \dots, n \tag{1}$$

where $(\dot{}) = \frac{d}{dt}()$. Here r_i is the growth rate of species i , a_{ij} (when $i \neq j$) accounts for the impact that species j has on the population change of species i , and the terms $a_{ii} x_i^2$ are adopted from Verhulst’s logistic growth model [35]. By collecting the individual populations $x_i(t)$ into a state vector $x(t) = [x_1(t), \dots, x_n(t)]^T$, Eq (1) can be represented in the compact form

$$\dot{x}(t) = \text{diag}(x(t))(r + Ax(t)), \tag{2}$$

where $r = [r_1, \dots, r_n]^T$ is a column vector of the growth rates, $A = (a_{ij})$ is the interspecific interaction matrix, and diag generates a diagonal matrix from a vector. Hereafter we drop the explicit time dependence of x .

Next we discuss the notion of fixed point, or equivalently steady state, in the GLV dynamics. This notion is important in the context of the human microbiome, as the measurements taken of the relative abundance of intestinal microbiota in the aforementioned studies typically represent steady behavior [4, 6]. In other words, the intestinal microbiota is a relatively resilient ecosystem [36, 37], and until the next large perturbation (e.g. antibiotic administration or dramatic change in diet) is introduced, the system remains stable for months and possibly even years [38–40]. The fixed points of system Eq (2) are those solutions x that satisfy $\dot{x} = 0$. The solution $x = 0$ (i.e. all species have zero abundance) is a trivial steady state. The set of non-trivial steady states contains those solutions x^* such that $r + Ax^* = 0$. When the matrix A is invertible, it follows that the non-trivial steady state $x^* = -A^{-1} r$ is unique [41].

Our study ultimately investigated the impact that different collections of microbial species have on their steady state abundances. In Fig 1 we presented a detailed analysis showing that if we introduce a new species into the ecosystem in Eq (2), the shift of the steady state is proportional to the interaction strengths between the newly introduced species and the previously

Impact of one new species: Consider the addition of one new species to the system described by (2), so that now we have $m = n + 1$ species, with the abundance of the i -th species denoted as $z_i(t)$, $i = 1, \dots, m$. The new m -dimensional state vector $z(t) = [z_1(t), z_2(t), \dots, z_m(t)]^T$ evolves as

$$\dot{z}(t) = \text{diag}(z(t))(g + Fz(t)), \tag{B1}$$

where

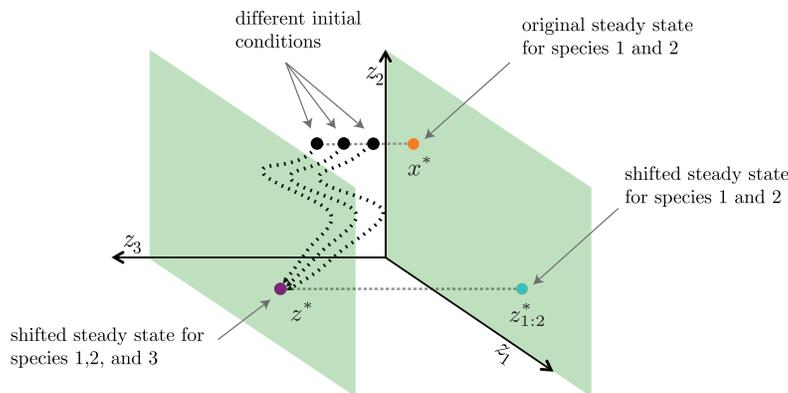
$$g = \begin{bmatrix} r \\ s \end{bmatrix}, \quad \text{and} \quad F = \begin{bmatrix} A & b \\ c^T & d \end{bmatrix}.$$

Note here that A and r are as defined in (2) and we have only introduced four new elements s, b, c, d . The scalar element s represents the growth rate of the additional species z_m . The n -dimensional vector b represents the impact that species z_m has on the first n species $z_{1:n}$, which also correspond to the n species in the original state vector x . The scalar element d is the Verhulst term for the new species z_m . Finally, the n -dimensional vector c represents the effect that $z_{1:n}$ have on the dynamics of the m -th state z_m .

Given the dynamics in (2) and (B1), with the assumption that the matrix A is full rank, we can show that the difference between the original steady state x^* and the shifted steady state of the same n species, denoted as $z_{1:n}^*$, satisfies the following equality

$$z_{1:n}^* - x^* = -A^{-1}bz_m^*, \tag{B2}$$

where z_m^* is the steady state value of the newly added species. Given that A is the same in (2) and (B1), the shift in the steady state of the original n species is bilinear in terms of the vector b and the steady state value of the newly added m -th species z_m^* .



Open-loop Control of Steady State: If there exists a diagonal matrix $P > 0$ such that $A^T P + PA < 0$ then for any z^* there exists b, c, d, s such that the original steady state x^* of (2) can be steered to z^* and furthermore this z^* can be made to be uniformly asymptotically stable (Supplementary Text Sec. 4.3.1, Theorem 8).

Impact of multiple non-common species: Consider two systems with different collections of species. Let \bar{z}^* be the steady state of system 1 and \hat{z}^* be the steady state of system 2. Assuming that the systems share n species, then we can just apply the results from (B2) recursively to calculate the difference between the abundances of the common species as

$$\bar{z}_{1:n}^* - \hat{z}_{1:n}^* = A^{-1} \left(\sum_{i \in \bar{M}} \bar{b}_i \bar{z}_i^* - \sum_{i \in \hat{M}} \hat{b}_i \hat{z}_i^* \right)$$

where \bar{M} and \hat{M} are the indices of the non-common species of system 1 and system 2 respectively.

Fig 1. Steady state shift in the generalized Lotka-Volterra model.

doi:10.1371/journal.pcbi.1004688.g001

existing ones. Similarly, if two communities with the same dynamics differ by only one species, then it is the interaction strength of that species with regard to the rest of the community that dictates how far apart the steady states of the two communities will be. This analytical result indicates that heterogeneity of interspecific interactions could lead to the clustering of steady states, and hence the emergence of community types.

To systematically investigate how changes in species collection affect the steady state shift in the GLV dynamics, we assumed that two microbial species will interact in the same fashion regardless of the host. Otherwise, if the interactions are host specific and the dynamics are classifiable, we can show that distinct community types will emerge almost trivially (S1 Text Sec. 6.2 and 7.1).

Metacommunity and Local Communities

Consider a universal species pool, also referred to as a metacommunity [42], indexed by a set of integers $\mathbf{S} = \{1, \dots, n\}$, an $n \times n$ matrix \mathbf{A} representing all possible pairwise interactions between species, and a vector \mathbf{r} of size n containing the growth rates for all the n species. The global parameters for the metacommunity are completely defined by the triple $(\mathbf{S}, \mathbf{A}, \mathbf{r})$. We consider q Local Communities (LCs), defined by sets $S^{[v]}$ that are subsets of \mathbf{S} , denoting the species present in LC_v , with $v = 1, \dots, q$. This modeling procedure is inspired by the fact that alternative community assembly scenarios could give rise to the compositional variations observed in the human microbiome [42]. These LCs represent microbial communities in the same body site across different subjects. For simplicity, we assume that each LC contains only p species ($p \leq n$), randomly selected from the metacommunity.

The GLV dynamics for each LC is given by

$$LC_v : \quad \dot{x}^{[v]}(t) = \text{diag}(x^{[v]}(t))(r^{[v]} + A^{[v]}x^{[v]}(t)), \tag{3}$$

where the LC specific interaction matrix and growth vector are defined as $A^{[v]} = \mathbf{A}_{S^{[v]}, S^{[v]}}$ and $r^{[v]} = \mathbf{r}_{S^{[v]}}$, respectively. That is, $A^{[v]}$ is obtained from \mathbf{A} by only taking the rows and columns of \mathbf{A} that are contained in the set $S^{[v]}$. A similar procedure is performed in order to obtain $r^{[v]}$. Finally for each $x^{[v]}$ there is a corresponding $\mathbf{x}^{[v]} \in \mathbb{R}^n$ that has the abundances for species $S^{[v]}$ of LC_v in the context of the metacommunity species pool \mathbf{S} .

To reveal the origins of community types in the human microbiome, we decomposed the universal interaction matrix as

$$\mathbf{A} = \mathbf{NH} \circ \mathbf{G}s, \tag{4}$$

which contains four components. (i) $\mathbf{N} \in \mathbb{R}^{n \times n}$ is the nominal interspecific interaction matrix where each element is sampled from a normal distribution with mean 0 and variance σ^2 , i.e. $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, \sigma^2)$. (ii) $\mathbf{H} \in \mathbb{R}^{n \times n}$ is a diagonal matrix that captures the overall interaction strength heterogeneity of different species. When studying the impact of interaction strength heterogeneity the diagonal elements of \mathbf{H} will be drawn from a power-law distribution with exponent $-\alpha$, i.e. $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$, which are subsequently normalized so that the mean of the diagonal elements is equal to 1. This is to ensure that the average interaction strength is bounded. For studies that do not involve interaction strength heterogeneity \mathbf{H} is simply the identity matrix. (iii) $\mathbf{G} \in \mathbb{R}^{n \times n}$ is the adjacency matrix of the underlying ecological network: $[\mathbf{G}]_{ij} = 1$ if species i is affected by the presence of species j and 0 otherwise. For details on the construction of \mathbf{G} for different network topologies see S1 Text Sec. 3.2.2. Note that the Hadamard product (\circ) between \mathbf{H} and \mathbf{G} represents element-wise matrix multiplication. (iv) The last component s is simply a scaling factor between 0 and 1. Finally, we set $[\mathbf{A}]_{ii} = -1$. The presence of the scaling factor s and setting the diagonal elements of \mathbf{A} to -1 are to ensure an asymptotic

stability condition for the GLV dynamics (S1 Text Sec. 4.2, 4.3.3, and 4.5). The elements in the global growth rate vector \mathbf{r} are taken from the uniform distribution, $[\mathbf{r}]_i \sim \mathcal{U}(0, 1)$. Details concerning the distribution \mathcal{N} , \mathcal{P} and \mathcal{U} can be found in S1 Text Sec. 3.1.1.

Origins of Community Types

We first studied the role of interspecific interaction strength heterogeneity on the emergence of community types. In order to achieve this, we chose the complete graph topology, i.e. each species interacts with all other species. This eliminates any structural heterogeneity. The nominal interaction strengths were taken from a normal distribution $\mathcal{N}(0, 1)$, the scaling component was set to $s = 0.7$, and the interaction strength heterogeneity was varied from low heterogeneity

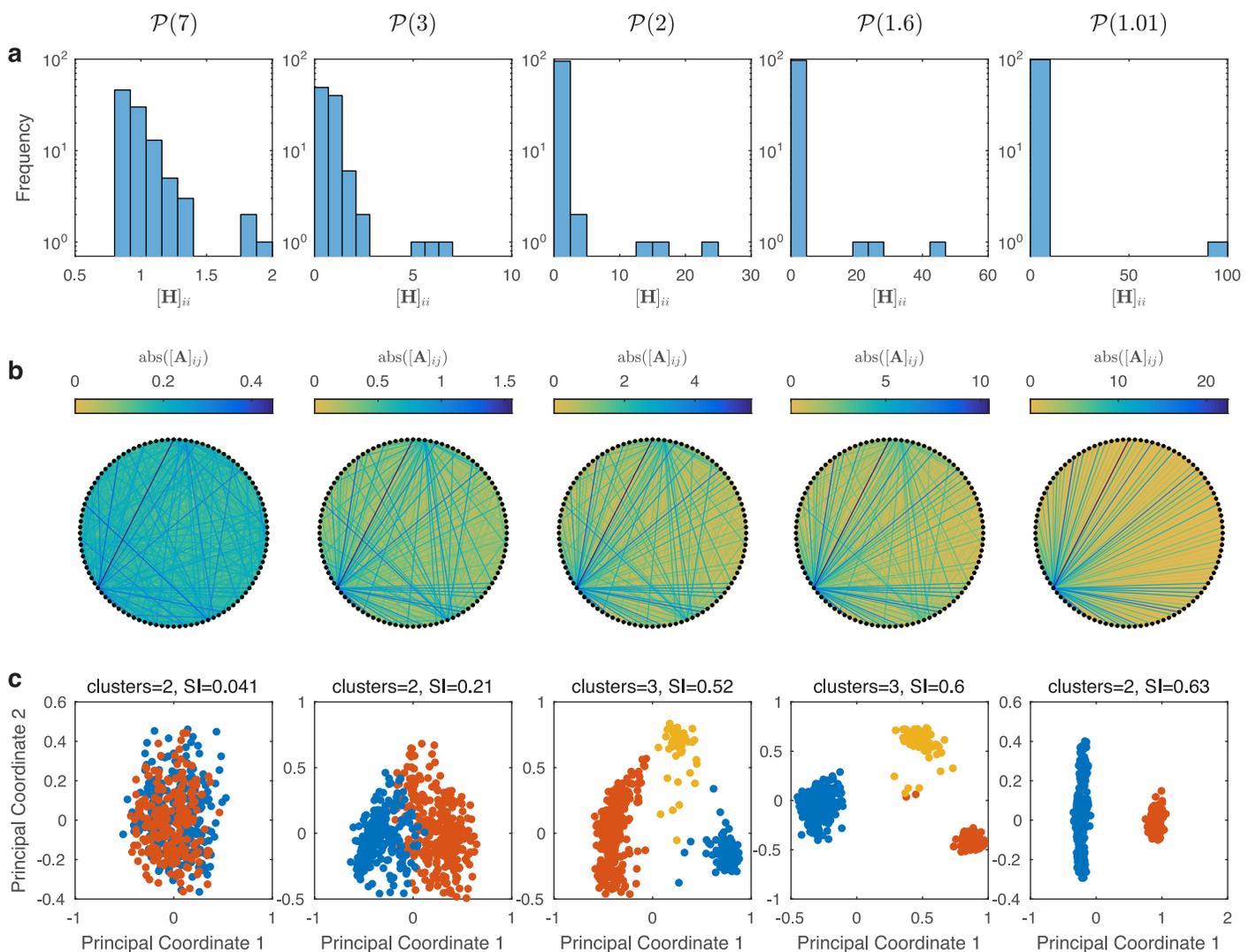


Fig 2. Impact of interaction strength heterogeneity on the distinctness of community types. A total of $q = 500$ local communities, each with $p = 80$ species randomly drawn from a universal pool of $n = 100$ species. The nominal components were drawn from $\mathcal{N}(0, 1)$, the interaction heterogeneity matrix elements were taken from $\mathcal{P}(\alpha)$ and α is varied with the set of values $\{7, 3, 2, 1.6, 1.01\}$ for each column in the figure. The topology component \mathbf{G} has all elements equal to 1, giving a complete graph. The scaling factor was set at $s = 0.07$. (a) Histogram of the diagonal elements of the heterogeneity matrix \mathbf{H} . (b) Visualization of the universal interaction matrix \mathbf{A} as a weighted adjacency matrix of a digraph. (c) Principle coordinate analyses of the normalized steady state for each local community using the Jensen-Shannon distance. The Silhouette Index and optimal number of clusters are denoted. Further details can be found in Materials and Methods.

doi:10.1371/journal.pcbi.1004688.g002

($\alpha = 7$) to a high level of heterogeneity ($\alpha = 1.01$). Fig 2 displays the distributions of the diagonal elements of the interaction heterogeneity matrix \mathbf{H} at various heterogeneity levels. For each level of heterogeneity we constructed 500 LCs, each with 80 species randomly drawn from a metacommunity of 100 species. Fig 2b illustrates the global interaction matrix \mathbf{A} as a weighted network. With low heterogeneity all the link weights are of the same order of magnitude. As the heterogeneity increases fewer nodes contain highly weighted links, until there is only one node with highly weighted links when $\alpha = 1.01$. These nodes with highly weighted links correspond to SISs.

Fig 2c presents the results of *Principle Coordinates Analysis* (PCoA) of the steady states associated with the 500 different LCs as a function of α . For low interaction heterogeneity ($\alpha = 7$) the classical clustering measure, Silhouette Index, is less than 0.1, suggesting a lack of clustering in the data. As the heterogeneity increases the steady states can be seen to separate in the first two principle coordinate axes. At one point ($\alpha = 2.0$) three clusters is the optimal number of clusters. Then as α continues to decrease the optimal number of clusters becomes two. The fact that there are three clusters when $\alpha = 2.0$ is not special, as a different number of optimal clusters can be observed with different model parameters or different clustering measures (see S1 Text Sec. 7.2) [7]. While the precise number of clusters is not important here, what is important is the fact that the degree of interaction strength heterogeneity controls the degree to which the clusters appear to be distinct. For low levels of interaction strength heterogeneity the clusters appear to be more like dense regions within a continuous gradient. As the heterogeneity increases, the clusters become more distinct. Indeed, having two clusters for $\alpha = 1.01$ is to be expected, because one of the clusters is associated with all the LCs that contain the single SIS, and the other LCs that do not contain the single SIS constitute the other cluster.

The overall trend observed in Fig 2c is unaffected if the complete graph is replaced by an *Erdős-Rényi* (ER) random graph, or if the total number of LCs is increased (S1 and S2 Figs). The result is also generally unaffected by the specifics of the nominal distribution (S1 Text Sec. 7.2.1), the mean degree of the ER graph (S1 Text Sec. 7.2.2), or the number of species in the LCs (S1 Text Sec. 7.2.3). Of course, each LC can be invaded by other species that are currently absent. If this migration occurs relatively fast, then all LCs will converge to roughly the same species collection and the clustering will disappear. Hence in our modeling approach we have to assume that the migration occurs at a relatively slow time scale, and the time interval between species invasions is too long to disrupt the clustering. We also note that if heterogeneous interactions are placed at random in the network the clustering of steady states does not arise (S3 Fig). Our results are also robust (in the control theoretical sense) to stochasticity and the migration of existing species [43]. Robustness to migration is illustrated in S4 and S5 Figs, and robustness to stochastic disturbances is illustrated in S6–S8 Figs (see S1 Text Sec. 4.4 for analytical robustness results).

We can explain the above results as follows: for low interaction strength heterogeneity all of the matrices $A^{[v]}$ are very similar. In other words, despite containing different sets of species, all the LCs have very similar dynamics. Thus, clustering of steady states is not to be expected. As the heterogeneity of interaction strength increases, however, some of the LCs will have species that are associated with the highly weighted columns in \mathbf{A} , i.e. the SISs. Fig 3 presents a detailed analysis of the most abundant (dominating) species in each of the three clusters (community types) in Fig 2c for $\alpha = 2$ and $\alpha = 1.6$, along with the abundances of the SISs within each cluster. It is clear that for different clusters their dominating species are different, consistent with the empirical finding that each enterotype is dominated by a different genus [6]. The SISs that are present in each cluster also vary. For instance with $\alpha = 1.6$ all LCs in the blue cluster contain SISs number 23 and 81, and none have species 60 or 51. For the orange cluster it is the opposite scenario. All of the LCs in the orange cluster contain SISs 60 and 51, and do not

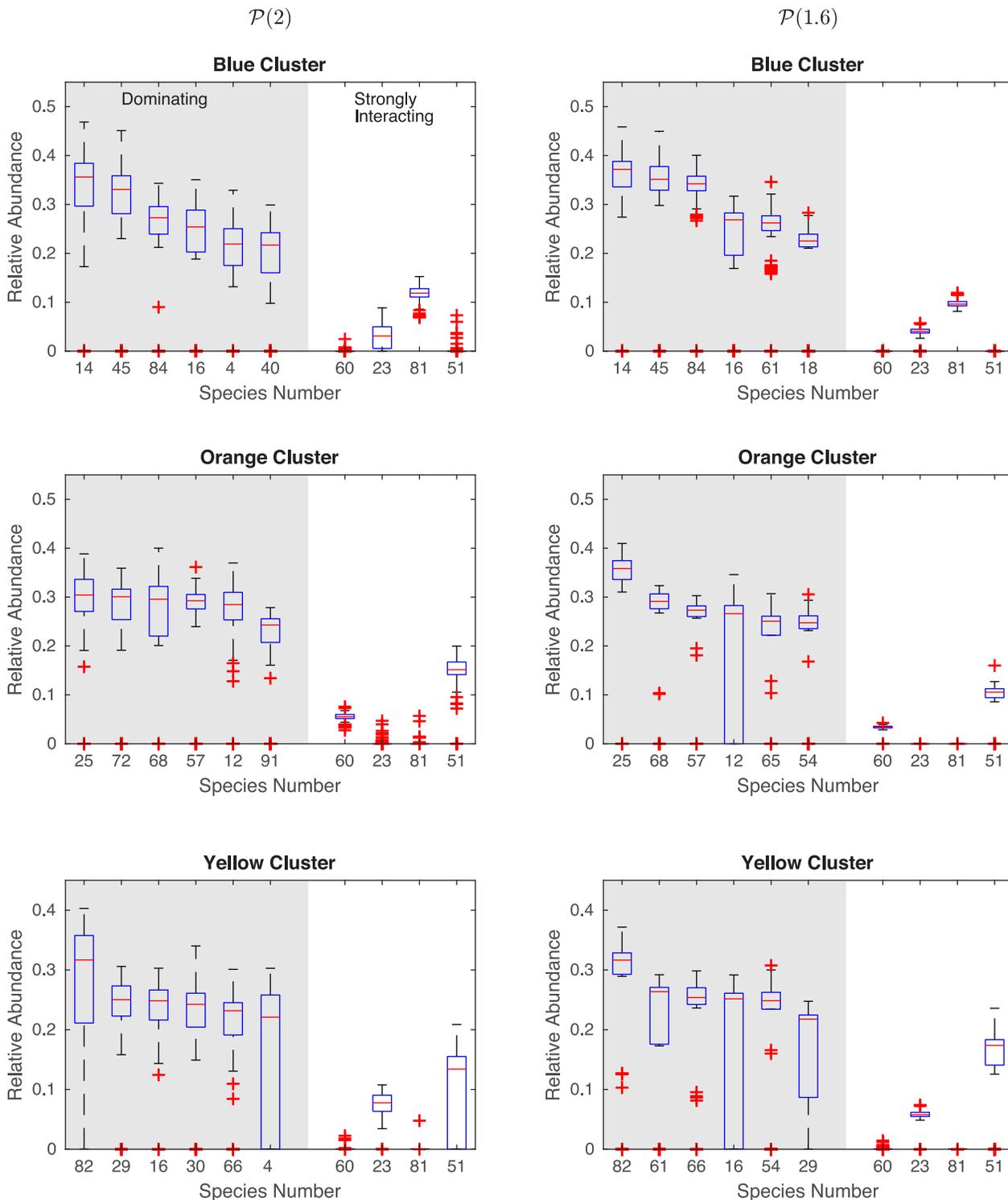


Fig 3. Comparison of dominating species to SISs in different community types (clusters). The relative abundances of the six most abundant species from each of the three clusters in Fig 2c for $\alpha = 2$ and $\alpha = 1.6$ are compared to that of the four species with the largest interaction strengths (60, 23, 81, and 51).

doi:10.1371/journal.pcbi.1004688.g003

contain species 23 or 81. Most of the LCs in the yellow cluster contain SISs 23 and 51. Hence, each community type is well characterized by a unique combination of SISs. Note that none of the SISs are dominating species. These findings, along with the analysis in Fig 1, suggest that heterogeneity in interaction strengths or the presence of SISs leads to the clustering of steady states, i.e. the emergence of community types.

We then studied the impact of structural heterogeneity on community types. Four different scenarios are illustrated in Fig 4: (a) a complete graph topology as in Fig 2; (b) an ER random graph as in S1 Fig; (c) a power-law out-degree network; (d) a power-law out-degree network with no interaction strength heterogeneity. Fig 4a, 4b and 4c support the main result shown in Fig 2, i.e. increasing interaction strength heterogeneity leads to the emergence of distinct community types. Fig 4d displays rather unexpected results as it suggests that structural heterogeneity alone does not lead to distinct community types. It is only with the inclusion of interaction strength heterogeneity that structurally heterogeneous microbial ecosystems can display strong clustering in their steady states as shown in Fig 4c. This result is rather surprising, because structural heterogeneity is observed in many real-world complex networks [44–46] and has been shown to affect many dynamical processes over complex networks [47–49].

Note that in the preparation of Fig 4 the steady state abundances were normalized to get relative abundances of the species and the Jensen-Shannon distance metric was used for clustering analysis [50]. The trends discussed above also hold when, instead of the Silhouette Index, the Variance Ratio Criterion is used as the clustering measure, or the Euclidean distance is used for

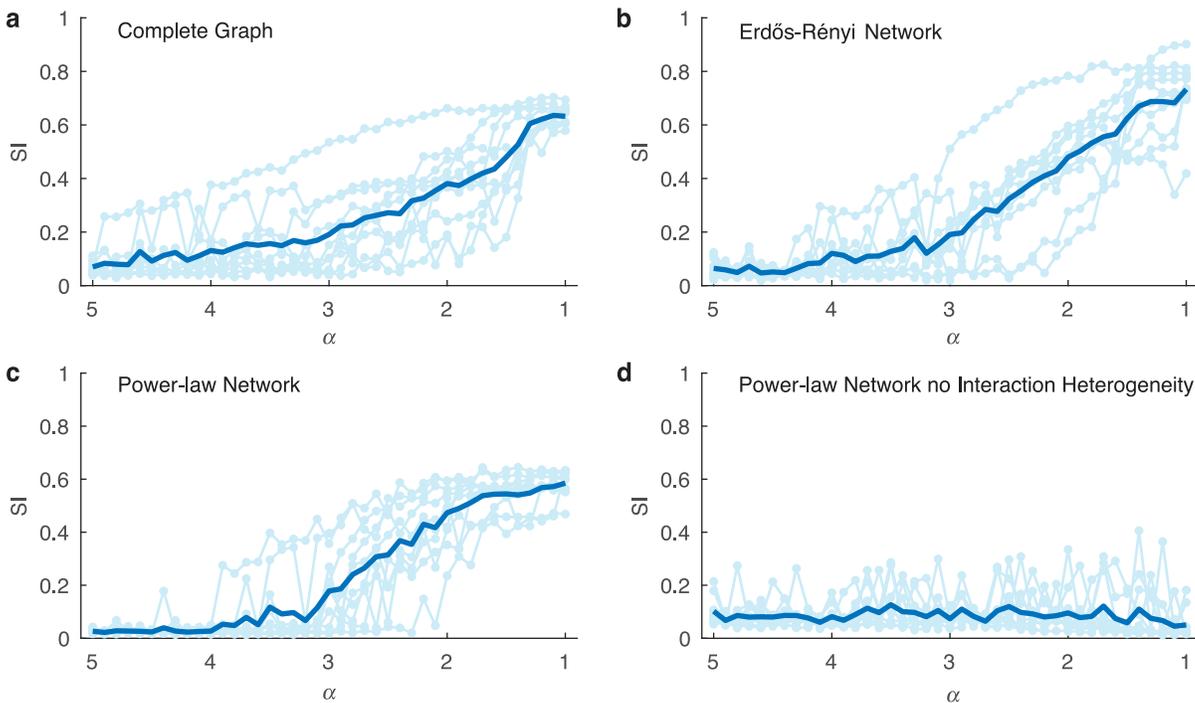


Fig 4. Impact of network structure on the distinctness of community types. For each type of network structure 10 different Universal Triples ($\mathbf{S}, \mathbf{A}, \mathbf{r}$) with $n = 100$ species and $q = 500$ local communities of size $p = 80$ were generated with results shown in the lighter color and averaged results shown in bold. (a) Complete graph. Same study as in Fig 2 with $\alpha \in [5, 1)$. (b) Erdős-Rényi network (digraph) $[\mathbf{N}]_{ij} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$, $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$ where $\alpha \in (1, 5]$, Probability $[\mathbf{G}]_{ij} = 1$ is 0.1, i.e. a mean in(out)-degree of 10, and scaling factor $s = 1/\sqrt{10}$. (c) Power-law out-degree network $[\mathbf{N}]_{ij} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$, $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$, \mathbf{G} is the adjacency matrix for a digraph with out-degree having a power-law distribution $\mathcal{P}(\alpha)$. The high-degree nodes have the largest interaction scaling. (d) Power-law out-degree network, no interactions strength heterogeneity $[\mathbf{N}]_{ij} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$, \mathbf{H} is the identity matrix, \mathbf{G} is the adjacency matrix for a digraph with out-degree having a distribution $\mathcal{P}(\alpha)$. Further details can be found in Materials and Methods.

doi:10.1371/journal.pcbi.1004688.g004

clustering, or when absolute abundances are analyzed along with the Euclidean distance being used (S9, S10 and S11 Figs). S11 Fig correlates to the analytical results in Fig 1, where absolute abundances and the Euclidean distance are implicitly used.

Control of Community Types

With the knowledge that each community type can be associated with a specific collection of SISs, we tested the hypothesis that a local community could be steered to a desired community type by controlling the combination of SISs only. Our results for three different scenarios are shown in Fig 5a for $\alpha = 1.6$. The local community that was controlled in each scenario is shown in magenta and is denoted LC^* , which initially belongs to the blue cluster. For Scenario 1, LC^* had the SISs 23 and 81 removed, with species 60 and 51 simultaneously introduced with random initial abundances drawn from $\mathcal{U}(0, 1)$. Recall that species 60 and 51 are the SISs present in the orange cluster. This swap of SISs shifts LC^* to a slightly different state (green dot) within the blue cluster. The GLV dynamics were then simulated and the trajectory goes from the blue cluster to the orange cluster. This result was independent of the initial condition of species 60 and 51 (Fig 5b). This open-loop control of the community type by manipulating a set of SISs also works at lower levels of heterogeneity (Fig 5c and 5d). Here we use the term open-loop to contrast closed-loop control where inputs are designed with feedback so as to continuously correct the system of interest. These findings imply that the SISs, despite their low abundances, can be used to effectively control a microbial community to a desired community type.

In Scenario 2 we tested if the same result could be obtained by removing the six most abundant species from LC^* and introducing the six most abundant species from the orange cluster at exactly the same abundance level as an arbitrary local community in the orange cluster. The state after this dominating species swap (red dot) starts close to the orange cluster, because the six most abundant species from a local community in that cluster were copied. The trajectory does not ultimately converge near the orange cluster, but goes toward the blue cluster instead. The trajectory, however, does not ultimately converge in the blue cluster because it does not contain any of the most abundant species present in the blue cluster.

In scenario 3 we explored how the open-loop control methodology just presented could also be used to conceptually justify the success of FMT in treating patients with rCDI [20–22]. This scenario begins by removing 20 species from LC^* (the top two SISs and 18 of the most abundant species) so as to emulate the effect of broad-spectrum antibiotics, resulting in an altered community (blue dot). Then the GLV dynamics were simulated and the local community converged to a new steady state (black dot), representing the CDI state. To emulate an oral capsule FMT 1% of the species abundances from an arbitrary LC in the orange cluster, i.e. the donor, was added to the CDI state, resulting in a slightly altered community (gray dot). The GLV dynamics were then simulated until the final steady state was reached (white dot). As expected the post-FMT steady state is in the orange cluster, the same cluster that is associated with the donor's LC. Note that if during the FMT the SISs in the donor's LC were not transplanted then the patient's post-FMT steady state does not converge in the orange cluster (S12 Fig).

The above results indicate that the presence of SISs simplifies the open-loop control design. However, the existence of community types is not a prerequisite for deploying this control methodology. The possibility for open-loop control of the human microbiome will likely be body site specific. Our work focused on the gut specifically because of the fact that this microbial community is very likely dominated by microbe-microbe and/or host-microbe interactions, rather than external disturbances. It is yet to be determined what factors drive the dynamics in other body sites.

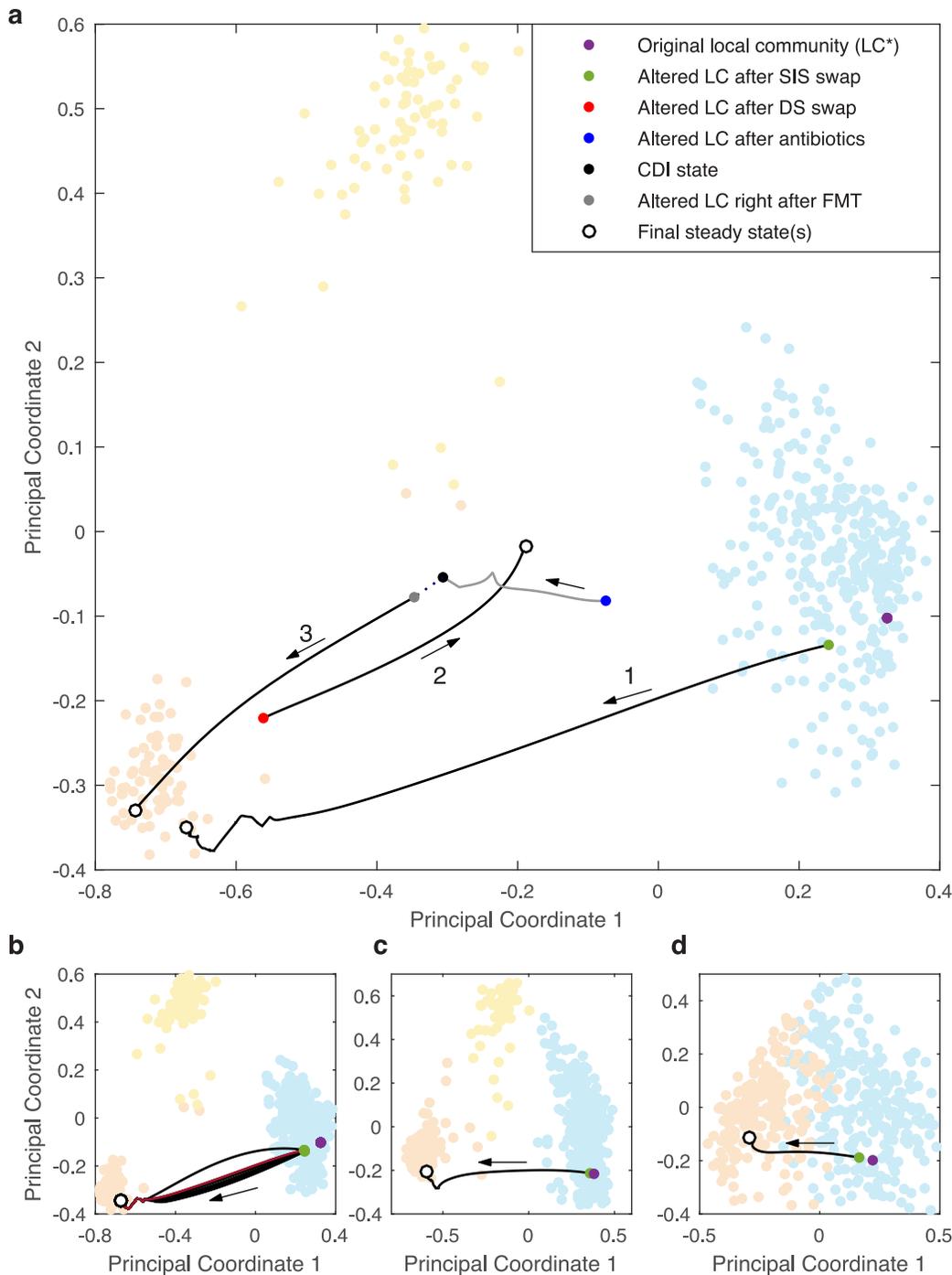


Fig 5. Open-loop control of the human microbiome. (a) Background of clustering analysis for $\alpha = 1.6$ from Fig 2, but with Euclidean distance used so that a projection matrix could be found to show the trajectories in the 2D principle coordinate plane (S1 Text Sec. 5.6). We aim to steer a local community (denoted as LC*, shown in magenta) in the blue cluster to the orange cluster. Three different scenarios are presented, per the three numbers above the arrows. Scenario 1: SISs swap. The SISs (23 and 81) of LC* were replaced by the SISs present in the orange cluster (60 and 51). The initial abundances of species 60 and 51 were drawn from $\mathcal{L}(0, 1)$, resulting in an altered community (green dot), and the GLV dynamics were simulated until the steady state was reached (white dot), which is located in the orange cluster as desired. Scenario 2: Dominating Species (DS) swap. The six most abundant species in LC* were removed and replaced by the six most abundant species from a local community in the orange cluster, with the initial condition after the switch of species shown as the red dot, and the dynamics were simulated until steady state was reached (white dot). Scenario 3: Fecal Microbiota Transplantation (FMT). The two SISs and 18 of the most abundant species (for a total of 20) were removed from LC* with the initial condition shown in blue (post-antibiotic state). Then the GLV dynamics were simulated (gray line) and the system converged to the black dot (CDI state). Then 1% of the steady abundances from an arbitrary LC in the orange cluster were added to the CDI state (gray dot, emulating oral capsule FMT) and the dynamics were then simulated until steady

state was reached. (b) The SISs swap process was repeated ten times, each time the initial abundances of species 60 and 51 were randomly drawn from $\mathcal{U}(0, 1)$. Nine of the simulations are shown in black and the simulation that pertains to Fig 4a is shown in maroon. (c) The same analysis as for Fig 5a, in terms of SISs swap, but for $\alpha = 2$. (d) The same analysis as for Fig 5a, in terms of SISs swap, but for $\alpha = 3$.

doi:10.1371/journal.pcbi.1004688.g005

Discussion

In this work we studied compositional shift as a function of species collection using a dynamic systems approach, aiming to offer a possible mechanism for the origins of community types. We found that the presence of interaction strength heterogeneity or SISs is sufficient to explain the emergence of community types in the human microbiome, independent of the topology of the underlying ecological network. The presence of heterogeneity in the interspecific interaction strengths in natural communities has been well studied in macroecology [23–25, 51]. Extensive studies are still required to explore this interesting direction in the human microbiome. While preliminary analysis is promising, all existing temporal metagenomic datasets are simply not sufficiently rich to infer the interspecific interaction strengths among all of the microbes present in and on our bodies [15] even at the genus level, let alone the species level. Recent studies have tried to overcome this issue by only investigating the interactions between the most abundant species [34]. Our results, however, suggest that SISs need not be the most abundant ones and can still play an important role in shaping the steady states of microbial ecosystems. Ignoring the lack of sufficient richness, system identification analysis with regularization and cross-validation [32, 52] of the largest temporal metagenomic dataset to date [39] does not disprove the existence of SISs. To the contrary, it supports this assertion (see S13 Fig). Permutation of the time series however also results in the identification of interaction strength heterogeneity (see S14 and S15 Figs). Hence, the presence of SISs needs to be systematically studied with novel system identification methods and perhaps further validated with co-culture experiments [15]. For example, we could first use metabolic network models to predict levels of competition and complementarity among species [53], which could then be used as prior information to further improve system identification [54].

Note that our notion of SIS is fundamentally different from that of keystone species, which are typically understood as species that have a disproportionately deleterious effect (relative to their abundance) on the community upon their removal [55]. One can apply a brute-force leave-one-out strategy to evaluate the “degree of keystone-ness” of any species in a given community [56]. Even without any interaction strength heterogeneity, a given community may still have a few keystone species. The SISs defined in this work are those species that have very strong impacts (either positive or negative) on the species that they directly interact with. The presence of SISs requires the presence of interaction strength heterogeneity. We emphasize that an SIS is not necessarily a keystone species. In fact, without any special structure embedded in the interaction matrix (and hence the ecological network), there is no reason why the removal of any SIS would cause a mass extinction. It does have a profound impact on the steady-state shift, which is exactly what we expected from our analytical results presented in Fig 1.

Our findings also have important implications as we move forward with developing microbiome-based therapies, whether it be through drastic diet changes, FMT, drugs, or even engineered microbes [57–63]. Indeed, our results suggest that a few strongly interacting microbes can determine the steady state landscape of the whole microbial community. Therefore, it may be possible to control the microbiome efficiently by controlling the collection of SISs present in a patient’s gut. Finer control may be possible through the engineering of microbes. This will involve a detailed mechanistic understanding of the metabolic pathways associated with the microbes of interest. As discussed in Fig 1, given a new steady state of interest, the parameters

b, c, d, s could be chosen such that the new steady state is feasible and stable (S1 Text Sec. 4.3.1). Then, with the knowledge of the appropriate parameters b, c, d, s it would be possible to introduce a known microbe with those characteristics or engineer one to have the desired properties. We emphasize that the stability and control of the microbial ecosystem must be studied at the macroscopic scale using a systems and control theoretic approach. This is similar to what is carried out in aerospace applications. The design of wings and control surfaces for an aircraft incorporate sophisticated fluid dynamic models. The control algorithms for planes however are often derived from simple linearized reduced order dynamic models where linear control techniques can be easily deployed [64]. Taken together, our results indicate that the origins and control of community types in the human microbiome can be explored analytically if we combine the tools of dynamic systems and control theory, opening new avenues to translational applications of the human microbiome.

Materials and Methods

The methods section begins with a toy example to illustrate the construction of the universal interaction matrix $\mathbf{A} = \mathbf{NH} \circ \mathbf{Gs}$ in Eq (4), where

$$\begin{aligned}
 \text{steps :} \quad (i) \quad \mathbf{N} &= \begin{bmatrix} 0 & 0.2 & 0.4 & -0.1 \\ 0.7 & 0 & 0.3 & 0.4 \\ -0.1 & 0.7 & 0 & 0.1 \\ -0.3 & -0.2 & 0.4 & 0 \end{bmatrix} \\
 (ii) \quad \mathbf{H} &= \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0.4 \end{bmatrix} \\
 (iii) \quad \mathbf{G} &= \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\
 (iv) \quad s &= 1 \\
 (v) \quad [\mathbf{A}]_{ii} &= -1 \\
 \text{final result :} \quad \mathbf{A} &= \begin{bmatrix} -1 & 0.04 & 0.08 & -0.04 \\ 7 & -1 & 0.06 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 0 & 0.08 & -1 \end{bmatrix}
 \end{aligned}$$

Given that \mathbf{H} is diagonal, it scales the columns of \mathbf{N} . If one thinks of \mathbf{A} as the adjacency matrix of a digraph, then \mathbf{H} scales all of the edges leaving a node. Thus one can consider \mathbf{H} as controlling the interaction strength heterogeneity of \mathbf{A} . Given the Hadamard product between \mathbf{H} and

\mathbf{G} , the off-diagonal elements of \mathbf{G} that are zero will result in the corresponding off-diagonal elements of \mathbf{A} being zero as well.

In the first study (Fig 2), to explore the impact of interaction heterogeneity on steady state shift, we varied the exponent $-\alpha$ of the power-law distribution of $[\mathbf{H}]_{ji}$ to generate five different universal interaction matrices \mathbf{A} of dimension 100×100 . For each universal interaction matrix \mathbf{A} , the nominal component \mathbf{N} consists of independent and identically distributed elements sampled from a normal distribution $\mathcal{N}(0, 1)$. The topology for this study was a complete graph and thus all the elements in \mathbf{G} are equal to 1. The heterogeneity element \mathbf{H} is constructed in two steps. First, five different vectors $\bar{h}(\alpha) \in \mathbb{R}^{100}$ are constructed where each element is sampled from a power-law distribution $\mathcal{P}(\alpha)$ for $\alpha \in \{7, 3, 1.6, 1.2, 1.01\}$. Then, each of the $\bar{h}(\alpha)$ is normalized to have a mean of 1, $h = \bar{h}/\text{mean}(\bar{h})$. Finally the heterogeneity matrix is defined as $\mathbf{H} = \text{diag}(h)$. For this study $s = 0.07$, ensuring uniform asymptotic stability for the case of low heterogeneity (see S1 Text Theorem 17). The final step in the construction of \mathbf{A} is to set the diagonal elements to -1 .

For each α the following simulation steps were taken. There are a total of 100 species, $\mathbf{S} = \{1, 2, \dots, 100\}$, in the metacommunity, and each of the 500 local communities contains 80 species, randomly chosen from \mathbf{S} . The MATLAB command used to perform this step is `randperm`. The initial condition for each of the 500 local communities, $x^{[v]}(0)$, were sampled from $\mathcal{U}(0, 1)$. The dynamics were then simulated for 100 seconds using the MATLAB command `ode45`. If any of the 500 simulations crashed due to instability or if the norm of the terminal discrete time derivative was greater than 0.01 then that local community was excluded from the rest of the study. Those simulations that finished without crashing and with small terminal discrete time derivative were deemed steady. Less than 1% of simulations were deemed unstable in the preparation of Fig 2. It is worth noting that by constructing the dynamics as described above the abundance profiles for our synthetic data do not contain the heavy-tailed abundance profile that is observed in the HMP gut data [4].

The networks presented in the second row of Fig 2 were constructed by considering \mathbf{A} as the weighted adjacency matrix of the network. Note that arrows showing directionality and self loops were suppressed. The links were color coded in proportion to the absolute value of the entries in \mathbf{A} .

For the last row of Fig 2 a clustering analysis was performed. For each α the steady state abundances of the 500 local communities were normalized so that we have 500 synthetic microbial samples. Then k -medoids clustering was performed for $k \in \{1, 2, \dots, 10\}$ using the Jensen-Shannon distance metric (S1 Text Sec. 5.1). Silhouette analysis was performed to determine the optimal number of clusters and the clustering results were illustrated in the 2-dimensional principle coordinates plot. For S1 Fig the same steps as for the preparation of Fig 2 were performed, but with \mathbf{G} representing the adjacency matrix of an Erdős-Rényi digraph with mean degree of 20 (mean in-degree of 10 and mean out-degree of 10) and $s = 1/\sqrt{10}$. Details on the construction of an Erdős-Rényi digraph can be found in S1 Text Section 3.2.1. For S2 Fig the same steps as above were performed in Fig 2 but with $p = 5,000$ local communities.

Fig 4 is a macroscopic analysis of how network structure plays a role in the steady state shift with values of $\alpha \in (1, 5]$. For each topology ten different universal matrices \mathbf{A} were generated. Fig 4a shows the results of a complete graph and for each of the ten universal \mathbf{A} the same steps as in the preparation of Fig 2 were carried out. Fig 4b shows the result of an Erdős-Rényi random digraph topology and for each of the ten \mathbf{A} matrices the same steps as in the preparation of S2 Fig were carried out. Fig 4c shows results for networks with a power-law out-degree distribution with a mean out-degree of 10, where the out-degree sequence uses the same \bar{h} in the construction of \mathbf{H} . More information on the construction of \mathbf{G} for a power-law out-degree

network can be found in [S1 Text](#) Sec. 3.2.2. [Fig 4d](#) shows results for networks with a power-law out-degree distribution with mean out-degree of 10 and there is no interaction strength heterogeneity, i.e. \mathbf{H} is the identity matrix. For this study the Silhouette Index was constructed from normalized steady state data using the Jensen-Shannon distance. [S9 Fig](#) is the same as [Fig 4](#), but instead of the Silhouette Index, the variance ratio criterion is used with the Jensen-Shannon distance, from normalized steady state abundance ([S1 Text](#) Sec. 5.4). In [S10 Fig](#) the Silhouette Index is determined from the Euclidean distance with normalized steady state abundance. Finally, in [S11 Fig](#) the Silhouette Index is determined by the Euclidean norm with the absolute steady state abundance.

[Fig 5](#) contains a PCoA analysis of the results from [Fig 2](#), but with the Euclidean distance being used instead of the Jensen-Shannon distance, making PCoA equivalent to principle component analysis. This enables us to project the open-loop control trajectories into the principle coordinates ([S1 Text](#) Sec 5.6). This procedure was also used in the preparation of [S12 Fig](#).

[S13–S15](#) Figs contain system identification analyses for temporal gut microbiome data of two subjects [[39](#)]. The data is publicly available from the metagenomics analysis server MG-RAST:4457768.3-4459735.3 and can also be accessed (as we did) from Qiita (<http://qiita.ucsd.edu>) under study ID 550. The processed data was downloaded as biom file “67_otu_table.biom” (2014-11-17 13:18:50.591389). The *Operational Taxonomic Units* (OTUs) were then grouped from the genus level and up, depending on the availability of known classifications for OTUs, and converted to a txt file using MacQIIME version 1.9.0-20140227 with the command `summarize_taxa.py` with the options `-L 6 -a true`. Data was collected over 445 days with 336 fecal samples from Subject A and 131 fecal samples from Subject B. Details on the system identification algorithm are now given. The dynamics in [Eq \(2\)](#) can be approximated in discrete time as [[32](#)]

$$e_i(k) + \log(x_i(t_{k+1})) - \log(x_i(t_k)) = r_i + \sum_{j=1}^n a_{ij} x_j(t_k) \tag{5}$$

for $i = 1, 2, \dots, n$ where $k = 1, 2, \dots, N - 1$ is the sample index, N is the total number of samples, t_k is the time stamp of sample k , and e is an error term that arises because of the assumption that $x(t)$ is constant over each interval $t \in [t_k, t_{k+1})$. [Eq \(5\)](#) can be rewritten in terms of a regressor vector

$$\phi(k) = [1, x_1(t_k), x_2(t_k), \dots, x_n(t_k)]^T,$$

the parameter vector $\theta_i = [r_i, a_{i1}, a_{i2}, \dots, a_{in}]$ and the log difference $y_i(k) = \log(x_i(t_{k+1})) - \log(x_i(t_k))$ as

$$e_i(k) + y_i(k) = \theta_i \phi(k).$$

The identification problem can then be defined as finding the parameter matrix estimate $\hat{\Theta} = [\hat{\theta}_1^T, \hat{\theta}_2^T, \dots, \hat{\theta}_n^T]^T$ of the true parameter matrix $\Theta = [\theta_1^T, \theta_2^T, \dots, \theta_n^T]^T$. Letting

$$y(k) = [y_1(k), y_2(k), \dots, y_n(k)]^T$$

be the log difference vector for all species and $Y = [y(1), y(2), \dots, y(N - 1)]$ be the log difference matrix the system identification problem can be compactly presented as

$$\min_{\hat{\Theta}} \|Y - \hat{\Theta}\Phi\|_F^2 + \lambda \|\hat{\Theta}\|_F^2$$

where $\Phi = [\phi(1), \phi(2), \dots, \phi(N - 1)]$ is the regressor matrix, $\|\cdot\|_F$ denotes the Frobenius norm, $\lambda \geq 0$ is the Tikhonov regularization term [[65](#)]. The minimal solution to the above problem

can be given directly as

$$\arg \min_{\hat{\Theta}} (\|Y - \hat{\Theta}\Phi\|_F^2 + \lambda \|\hat{\Theta}\|_F^2) = Y\Phi^T(\Phi\Phi^T + \lambda I)^{-1}$$

where I is the identity matrix.

Next we discuss how missing data, zero reads, and λ were chosen. The difference equation in Eq (5) only uses sample data over two consecutive time samples. Therefore, in the construction of Y and Φ we only include samples that for which there is data from the next day as well. Also, given that logarithms are used, when a sample has zero reads for a given taxa, a read value of one is inserted. Then relative abundances are computed before the logarithm is taken. Finally we discuss how the regularization parameter is chosen. For S13 and S14 Figs the following cross-validation is performed. For Subjects A and B two-thirds of data was used for training and one-third for testing. More precisely, for each λ two-thirds of the data from Subject A and two-thirds of the data from Subject B were used to identify their corresponding dynamical constants. Then the combined error from the two test sets was used to find the optimal λ . The regularization value used in S15 Fig is simply the same regularization value used in S13 Fig.

Supporting Information

S1 Fig. Impact of interaction strength heterogeneity on the distinctness of community types. Same as Fig 2 but with the topology component G chosen to be an Erdős-Rényi digraph with a link probability of 0.1 and the scaling factor was set at $s = 1/\sqrt{10}$.
(EPS)

S2 Fig. Impact of interaction strength heterogeneity on the distinctness of community types. Same as Fig 2 but with $p = 5,000$ local communities. Note that it is rather counter-intuitive that for $\alpha = 1.01$ the Silhouette Index suggests that there are two clusters, while PCoA suggests three clusters. We emphasize that as a typical ordination method, the PCoA just produces a spatial representation of the entities in the dataset, rather than the actual determination of cluster membership. Note that as compared to Fig 2, because there are more samples in this figure, the distinctness of the clusters when $\alpha = 2$ has shifted to more of a continuous gradient as apposed to distinct clusters.
(EPS)

S3 Fig. Impact of interaction heterogeneity disbursed randomly throughout the network. The set up is the same as that of Fig 2 but instead of H being a diagonal matrix, it is a full matrix, so that individual interactions are scaled randomly from a power-law distribution.
(EPS)

S4 Fig. Impact of low levels of migration. Same as Fig 2 but with a new term $\lambda(t) \in [0, 1]^n$ added to the dynamics so that now $\dot{x} = \lambda + \text{diag}(x)(r + Ax)$. In this example $\lambda_i \sim \mathcal{U}(0, 0.1)$. The disturbance is sampled every 0.01 seconds and held constant until the next sample is taken.
(EPS)

S5 Fig. Impact of moderate levels of migration. Same as Fig 2 but with a new term $\lambda(t) \in [0, 1]^n$ added to the dynamics so that now $\dot{x} = \lambda(t) + \text{diag}(x)(r + Ax)$. In this example $\lambda_i \sim \mathcal{U}(0, 1)$. The disturbance is sampled every 0.01 seconds and held constant until the next sample is taken.
(EPS)

S6 Fig. Impact of small stochastic disturbance. Same as [Fig 2](#) but with stochastic Itô dynamics $dx = \text{diag}(x)(r dt + Ax dt + c dw)$ where w is a n -dimensional Brownian motion and c represents the stochastic disturbance strength. Dynamics were simulated with a discrete time step of 0.01 seconds and $c = 0.1$.

(EPS)

S7 Fig. Impact of moderate stochastic disturbance. Same as [Fig 2](#) but with stochastic Itô dynamics $dx = \text{diag}(x)(r dt + Ax dt + c dw)$ where w is a n -dimensional Brownian motion and c represents the stochastic disturbance strength. Dynamics were simulated with a discrete time step of 0.01 seconds and $c = 0.5$.

(EPS)

S8 Fig. Impact of large stochastic disturbance. Same as [Fig 2](#) but with stochastic Itô dynamics $dx = \text{diag}(x)(r dt + Ax dt + c dw)$ where w is a n -dimensional Brownian motion and c represents the stochastic disturbance strength. Dynamics were simulated with a discrete time step of 0.01 seconds and $c = 1$.

(EPS)

S9 Fig. Impact of network structure on the distinctness of community types. The same as [Fig 4](#) with the *Variance Ratio Criterion* (VRC) used as apposed to the Silhouette Index for the clustering measure. See [S1 Text](#) Sec. 5.4 for details on the VRC.

(EPS)

S10 Fig. Impact of network structure on the distinctness of community types. The same as [Fig 4](#) with the Euclidean distance metric used instead of the Jensen-Shannon distance metric.

(EPS)

S11 Fig. Impact of network structure on the distinctness of community types. The same as [Fig 4](#) with the Euclidean distance metric used instead of the Jensen-Shannon distance metric and absolute abundance used instead of relative abundance.

(EPS)

S12 Fig. Unsuccessful fecal microbiota transplantation. Similar to Scenario 3 shown in [Fig 5a](#), but during the FMT, the SISs (60 and 51) of the donor's local community in the orange cluster were not transplanted to the CDI state (black dot). This FMT resulted in a slightly altered community (gray dot) and the system eventually evolved to a steady state (white dot) that is not in the orange cluster. Hence the FMT failed.

(EPS)

S13 Fig. System identification, Tikhonov regularization $\lambda = 0.0423$. System identification was performed on the stool samples from the longitudinal data in [\[39\]](#) for two subjects as described in [S1 Text](#) where λ was determined by cross-validation. (a) Visualization of microbial taxa in terms of relative abundances versus day sample was taken. (b) Heat map of the interaction matrix for top 100 SISs. (c) Histogram of *Standard Deviation* (SD) of the columns of the interaction matrix. (d) List of top ten SISs in descending interaction strength (defined by the SD of each column in the interaction matrix) with relative abundances over all samples shown as a box plot. The banded structure shown in the heat map supports the assertion that SISs do exist in the gut microbiome. However this banded structure is also seen when the dates of the sample collections are permuted, see [S14](#) and [S15](#) Figs.

(EPS)

S14 Fig. System identification, day swap, Tikhonov regularization $\lambda = 0.0057$. System identification was performed on the stool samples from the longitudinal data in [\[39\]](#), but with the

collection dates permuted, λ was determined by cross-validation on the permuted data. (a) Visualization of microbial taxa in terms of relative abundances versus day sample was taken (not permuted samples). (b) Heat map of the interaction matrix for top 100 SISs. (c) Histogram of *Standard Deviation* (SD) of the columns of the interaction matrix. (d) List of top ten SISs in descending interaction strength (defined by the SD of each column in the interaction matrix) with relative abundances over all samples shown as a box plot. Even though the sample days have been permuted the banded structure still persists.
(EPS)

S15 Fig. System identification, day swap, Tikhonov regularization $\lambda = 0.0423$. System identification was performed on the stool samples from the longitudinal data in [39], but with the collection dates permuted, λ was selected to be the same as in S13 Fig. (a) Visualization of microbial taxa in terms of relative abundances versus day sample was taken (not permuted samples). (b) Heat map of the interaction matrix for top 100 SISs. (c) Histogram of *Standard Deviation* (SD) of the columns of the interaction matrix. (d) List of top ten SISs in descending interaction strength (defined by the SD of each column in the interaction matrix) with relative abundances over all samples shown as a box plot. For the permuted data when λ is larger than the optimal value from the cross-validation the identification method biases towards making the most abundant species also the SISs.
(EPS)

S1 Text. Detailed treatment of necessary mathematical components. S1 Text contains discussions on: random variables, random matrices, dynamic stability, clustering, ordination, modeling, and more simulation results.
(PDF)

Acknowledgments

Special thanks to Aimee Milliken for a careful reading of the text.

Author Contributions

Conceived and designed the experiments: YYL. Performed the experiments: TEG AB. Analyzed the data: TEG YYL AB HTC STW. Contributed reagents/materials/analysis tools: TEG HTC AB. Wrote the paper: TEG YYL. Edited the manuscript: AB STW.

References

1. DuPont AW, DuPont HL. The intestinal microbiota and chronic disorders of the gut. *Nat Rev Gastroenterol Hepatol*. 2011 09; 8(9):523–531. PMID: [21844910](#)
2. Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol*. 2009 05; 9(5):313–323. PMID: [19343057](#)
3. Nelson KE, Ehrlich SD. *MetaHIT: The European Union Project on Metagenomics of the Human Intestinal Tract*. Springer New York; 2011.
4. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012; 486(7402):207–214. doi: [10.1038/nature11234](#) PMID: [22699609](#)
5. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature*. 2012; 486(7402):215–221. doi: [10.1038/nature11209](#) PMID: [22699610](#)
6. Arumugam M, et al. Enterotypes of the human gut microbiome. *Nature*. 2011; 473(7346):174–180. PMID: [21508958](#)
7. Koren O, Knights D, Gonzalez A, Waldron L, Segata N, Knight R, et al. A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. *PLoS Comput Biol*. 2013; 9(1). PMID: [23326225](#)

8. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science*. 2011; 334(6052):105–108. doi: [10.1126/science.1208344](https://doi.org/10.1126/science.1208344) PMID: [21885731](https://pubmed.ncbi.nlm.nih.gov/21885731/)
9. Hildebrand F, Nguyen T, Brinkman B, Yunta RG, Cauwe B, Vandenabeele P, et al. Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biology*. 2013; doi: [10.1186/gb-2013-14-1-r4](https://doi.org/10.1186/gb-2013-14-1-r4) PMID: [23347395](https://pubmed.ncbi.nlm.nih.gov/23347395/)
10. Zhou Y, Mihindukulasuriya KA, Gao H, La Rosa PS, Wylie KM, Martin JC, et al. Exploration of bacterial community classes in major human habitats. *Genome Biology*. 2014; doi: [10.1186/gb-2014-15-5-r66](https://doi.org/10.1186/gb-2014-15-5-r66)
11. Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. *Nature*. 2014 05; 509(7500):357–360. PMID: [24739969](https://pubmed.ncbi.nlm.nih.gov/24739969/)
12. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences*. 2011; 108:4680–4687. doi: [10.1073/pnas.1002611107](https://doi.org/10.1073/pnas.1002611107)
13. Jeffery IB, Claesson MJ, O'Toole PW, Shanahan F. Categorization of the gut microbiota: enterotypes or gradients? *Nat Rev Micro*. 2012 09; 10(9):591–592.
14. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, et al. Gut microbiota composition correlates with diet and health in the elderly. *Nature*. 2012 08; 488(7410):178–184. PMID: [22797518](https://pubmed.ncbi.nlm.nih.gov/22797518/)
15. Faust K, Raes J. Microbial interactions: from networks to models. *Nature Reviews Microbiology*. 2012; 10(8):538–550. PMID: [22796884](https://pubmed.ncbi.nlm.nih.gov/22796884/)
16. Knights D, Ward TL, McKinlay CE, Miller H, Gonzalez A, McDonald D, et al. Rethinking “Enterotypes”. *Cell Host & Microbe*. 2014; 16(4):433–437.
17. Arumugam M, et al. Addendum: Enterotypes of the human gut microbiome. *Nature*. 2014; 506(7489): 516–516. doi: [10.1038/nature13075](https://doi.org/10.1038/nature13075)
18. Lewontin RC. The meaning of stability. In: *Brookhaven symposia in biology*. vol. 22; 1969. p. 13.
19. Connell JH, Sousa WP. On the evidence needed to judge ecological stability or persistence. *American Naturalist*. 1983; p. 789–824. doi: [10.1086/284105](https://doi.org/10.1086/284105)
20. Youngster I, Sauk J, Pindar C, Wilson RG, Kaplan JL, Smith MB, et al. Fecal Microbiota Transplant for Relapsing *Clostridium difficile* Infection Using a Frozen Inoculum From Unrelated Donors: A Randomized, Open-Label, Controlled Pilot Study. *Clinical Infectious Diseases*. 2014; 58(11):1515–1522. doi: [10.1093/cid/ciu135](https://doi.org/10.1093/cid/ciu135) PMID: [24762631](https://pubmed.ncbi.nlm.nih.gov/24762631/)
21. Weingarden A, González A, Vázquez-Baeza Y, Weiss S, Humphry G, Berg-Lyons D, et al. Dynamic changes in short-and long-term bacterial composition following fecal microbiota transplantation for recurrent *Clostridium difficile* infection. *Microbiome*. 2015; 3(1):10. PMID: [25825673](https://pubmed.ncbi.nlm.nih.gov/25825673/)
22. Youngster I, Russell G, Pindar C, Ziv-Baran T, Sauk J, Hohmann E. Oral, capsulized, frozen fecal microbiota transplantation for relapsing *clostridium difficile* infection. *JAMA*. 2014; 312(17):1772–1778. doi: [10.1001/jama.2014.13875](https://doi.org/10.1001/jama.2014.13875) PMID: [25322359](https://pubmed.ncbi.nlm.nih.gov/25322359/)
23. Paine RT. Food-web analysis through field measurement of per capita interaction strength. *Nature*. 1992 01; 355(6355):73–75. doi: [10.1038/355073a0](https://doi.org/10.1038/355073a0)
24. Eklöf A, Ebenman B. Species loss and secondary extinctions in simple and complex model communities. *Journal of Animal Ecology*. 2006; 75(1):239–246. doi: [10.1111/j.1365-2656.2006.01041.x](https://doi.org/10.1111/j.1365-2656.2006.01041.x) PMID: [16903061](https://pubmed.ncbi.nlm.nih.gov/16903061/)
25. Emmerson MC, Raffaelli D. Predator—prey body size, interaction strength and the stability of a real food web. *Journal of Animal Ecology*. 2004; 73(3):399–409.
26. Gerber GK. The dynamic microbiome. *FEBS Letters*. 2014 11; 588(22):4131–4139. doi: [10.1016/j.febslet.2014.02.037](https://doi.org/10.1016/j.febslet.2014.02.037) PMID: [24583074](https://pubmed.ncbi.nlm.nih.gov/24583074/)
27. Muñoz-Tamayo R, Laroche B, Walter É, Doré J, Leclerc M. Mathematical modelling of carbohydrate degradation by human colonic microbiota. *Journal of Theoretical Biology*. 2010 9; 266(1):189–201. doi: [10.1016/j.jtbi.2010.05.040](https://doi.org/10.1016/j.jtbi.2010.05.040) PMID: [20561534](https://pubmed.ncbi.nlm.nih.gov/20561534/)
28. Shoaie S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S, Nielsen J. Understanding the interactions between bacteria in the human gut through metabolic modeling. *Sci Rep*. 2013 08; 3. doi: [10.1038/srep02532](https://doi.org/10.1038/srep02532) PMID: [23982459](https://pubmed.ncbi.nlm.nih.gov/23982459/)
29. Waldram A, Holmes E, Wang Y, Rantalainen M, Wilson ID, Tuohy KM, et al. Top-down systems biology modeling of host metabotype- microbiome associations in obese rodents. *Journal of proteome research*. 2009; 8(5):2361–2375. PMID: [19275195](https://pubmed.ncbi.nlm.nih.gov/19275195/)
30. Bucci V, Xavier JB. Towards Predictive Models of the Human Gut Microbiome. *Journal of Molecular Biology*. 2014; 426(23):3907–3916. doi: [10.1016/j.jmb.2014.03.017](https://doi.org/10.1016/j.jmb.2014.03.017) PMID: [24727124](https://pubmed.ncbi.nlm.nih.gov/24727124/)

31. Pepper JW, Rosenfeld S. The emerging medical ecology of the human gut microbiome. *Trends in ecology & evolution*. 2012; 27(7):381–384.
32. Stein RR, Bucci V, Toussaint NC, Buffie CG, R atsch G, Pamer EG, et al. Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. *PLoS Comput Biol*. 2013; 9(12). doi: [10.1371/journal.pcbi.1003388](https://doi.org/10.1371/journal.pcbi.1003388) PMID: [24348232](https://pubmed.ncbi.nlm.nih.gov/24348232/)
33. Marino S, Baxter NT, Huffnagle GB, Petrosino JF, Schloss PD. Mathematical modeling of primary succession of murine intestinal microbiota. *Proceedings of the National Academy of Sciences*. 2014; 111(1):439–444. doi: [10.1073/pnas.1311322111](https://doi.org/10.1073/pnas.1311322111)
34. Fisher CK, Mehta P. Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries Using Sparse Linear Regression. *PLoS ONE*. 2014; 9(7):e102451. doi: [10.1371/journal.pone.0102451](https://doi.org/10.1371/journal.pone.0102451) PMID: [25054627](https://pubmed.ncbi.nlm.nih.gov/25054627/)
35. Goel NS, Maitra SC, Montroll EW. On the Volterra and other nonlinear models of interacting populations. *Reviews of modern physics*. 1971; 43(2):231. doi: [10.1103/RevModPhys.43.231](https://doi.org/10.1103/RevModPhys.43.231)
36. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012 09; 489(7415):220–230. PMID: [22972295](https://pubmed.ncbi.nlm.nih.gov/22972295/)
37. Relman DA. The human microbiome: ecosystem resilience and health. *Nutrition reviews*. 2012; 70(suppl 1):S2–S9. doi: [10.1111/j.1753-4887.2012.00489.x](https://doi.org/10.1111/j.1753-4887.2012.00489.x) PMID: [22861804](https://pubmed.ncbi.nlm.nih.gov/22861804/)
38. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biology*. 2014; doi: [10.1186/gb-2014-15-7-r89](https://doi.org/10.1186/gb-2014-15-7-r89) PMID: [25146375](https://pubmed.ncbi.nlm.nih.gov/25146375/)
39. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, et al. Moving pictures of the human microbiome. *Genome Biology*. 2011; doi: [10.1186/gb-2011-12-5-r50](https://doi.org/10.1186/gb-2011-12-5-r50) PMID: [21624126](https://pubmed.ncbi.nlm.nih.gov/21624126/)
40. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The long-term stability of the human gut microbiota. *Science*. 2013; 341(6141). doi: [10.1126/science.1237439](https://doi.org/10.1126/science.1237439)
41. Goh BS. Global stability in many-species systems. *American Naturalist*. 1977; p. 135–143. doi: [10.1086/283144](https://doi.org/10.1086/283144)
42. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. The Application of Ecological Theory Toward an Understanding of the Human Microbiome. *Science*. 2012; 336(6086):1255–1262. doi: [10.1126/science.1224203](https://doi.org/10.1126/science.1224203) PMID: [22674335](https://pubmed.ncbi.nlm.nih.gov/22674335/)
43. Fisher CK, Mehta P. The transition between the niche and neutral regimes in ecology. *Proceedings of the National Academy of Sciences*. 2014; 111(36):13111–13116. doi: [10.1073/pnas.1405637111](https://doi.org/10.1073/pnas.1405637111)
44. Barab asi AL, Albert R. Emergence of scaling in random networks. *science*. 1999; 286(5439):509–512. doi: [10.1126/science.286.5439.509](https://doi.org/10.1126/science.286.5439.509) PMID: [10521342](https://pubmed.ncbi.nlm.nih.gov/10521342/)
45. Li L, Alderson D, Doyle JC, Willinger W. Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics*. 2004; 2(4):431–523. doi: [10.1080/15427951.2005.10129111](https://doi.org/10.1080/15427951.2005.10129111)
46. Alderson DL, Doyle JC. Contrasting views of complexity and their implications for network-centric infrastructures. *Systems, Man and Cybernetics, Part A: Systems and Humans*, IEEE Transactions on. 2010; 40(4):839–852. doi: [10.1109/TSMCA.2010.2048027](https://doi.org/10.1109/TSMCA.2010.2048027)
47. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. *Physical review letters*. 2001; 86(14):3200. doi: [10.1103/PhysRevLett.86.3200](https://doi.org/10.1103/PhysRevLett.86.3200) PMID: [11290142](https://pubmed.ncbi.nlm.nih.gov/11290142/)
48. Nishikawa T, Motter AE, Lai YC, Hoppensteadt FC. Heterogeneity in oscillator networks: Are smaller worlds easier to synchronize? *Physical review letters*. 2003; 91(1):014101. doi: [10.1103/PhysRevLett.91.014101](https://doi.org/10.1103/PhysRevLett.91.014101) PMID: [12906539](https://pubmed.ncbi.nlm.nih.gov/12906539/)
49. Liu YY, Slotine JJ, Barab asi AL. Controllability of complex networks. *Nature*. 2011; 473(7346):167–173. PMID: [21562557](https://pubmed.ncbi.nlm.nih.gov/21562557/)
50. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, et al. Conducting a Microbiome Study. *Cell*. 2014 7; 158(2):250–262. PMID: [25036628](https://pubmed.ncbi.nlm.nih.gov/25036628/)
51. McCann K, Hastings A, Huxel GR. Weak trophic interactions and the balance of nature. *Nature*. 1998 10; 395(6704):794–798.
52. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gouberne A, et al. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature*. 2015; 517(7533):205–208. PMID: [25337874](https://pubmed.ncbi.nlm.nih.gov/25337874/)
53. Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences*. 2013; 110(31):12804–12809. doi: [10.1073/pnas.1300926110](https://doi.org/10.1073/pnas.1300926110)
54. Angulo MT, Moreno JA, Barab asi AL, Liu YY. Fundamental limitations of network reconstruction. *arXiv*. 2015; Available from: <http://arxiv.org/abs/1508.03559>.
55. Paine RT. A conversation on refining the concept of keystone species. *JSTOR*; 1995.

56. Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in microbiology*. 2014; 5. doi: [10.3389/fmicb.2014.00219](https://doi.org/10.3389/fmicb.2014.00219) PMID: [24904535](https://pubmed.ncbi.nlm.nih.gov/24904535/)
57. Sonnenburg JL. Engineering the Human Microbiome Shows Promise for Treating Disease. *Scientific American*. 2015; 312(3).
58. Yaung SJ, Church GM, Wang HH. Recent Progress in Engineering Human-Associated Microbiomes. In: *Engineering and Analyzing Multicellular Systems*. Springer; 2014. p. 3–25.
59. Tanouchi Y, Smith RP, You L. Engineering microbial systems to explore ecological and evolutionary dynamics. *Current opinion in biotechnology*. 2012; 23(5):791–797. doi: [10.1016/j.copbio.2012.01.006](https://doi.org/10.1016/j.copbio.2012.01.006) PMID: [22310174](https://pubmed.ncbi.nlm.nih.gov/22310174/)
60. Leonard E, Nielsen D, Solomon K, Prather KJ. Engineering microbes with synthetic biology frameworks. *Trends in biotechnology*. 2008; 26(12):674–681. doi: [10.1016/j.tibtech.2008.08.003](https://doi.org/10.1016/j.tibtech.2008.08.003) PMID: [18977048](https://pubmed.ncbi.nlm.nih.gov/18977048/)
61. Yaung SJ, Deng L, Li N, Braff JL, Church GM, Bry L, et al. Improving microbial fitness in the mammalian gut by in vivo temporal functional metagenomics. *Molecular systems biology*. 2015; 11(3). doi: [10.15252/msb.20145866](https://doi.org/10.15252/msb.20145866)
62. Mee MT, Collins JJ, Church GM, Wang HH. Syntrophic exchange in synthetic microbial communities. *Proceedings of the National Academy of Sciences*. 2014; 111(20):E2149–E2156. doi: [10.1073/pnas.1405641111](https://doi.org/10.1073/pnas.1405641111)
63. Kotula JW, Kerns SJ, Shaket LA, Siraj L, Collins JJ, Way JC, et al. Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proceedings of the National Academy of Sciences*. 2014; 111(13):4838–4843. doi: [10.1073/pnas.1321321111](https://doi.org/10.1073/pnas.1321321111)
64. Stevens BL, Lewis FL. *Aircraft Control and Simulation*. Wiley; 2003.
65. Tikhonov A. Solution of incorrectly formulated problems and the regularization method. In: *Soviet Math. Dokl.* vol. 5; 1963. p. 1035–1038.